(54) Title: CAMERA MOTION PARAMETERS ESTIMATION METHOD
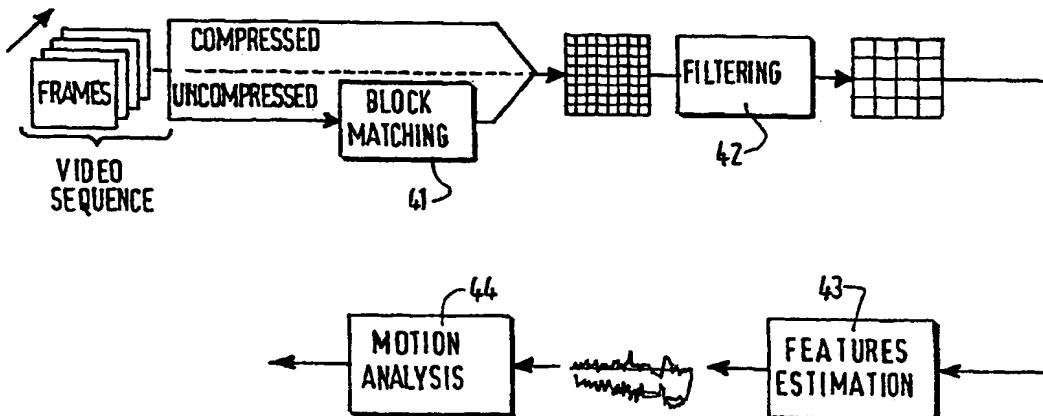
(57) Abstract

The invention relates to a method of estimation of the motion parameters of a camera. It is applied to a sequence of successive video frames subdivided into blocks and comprises the following steps of extracting from the sequence vectors corresponding to the motion between two frames said motion vectors forming the camera velocity field, preprocessing the camera velocity field thus obtained, in order to reduce the amount of data and the heterogeneousness of the extracted motion vectors, estimating for each pair of frames, from said preprocessed field, camera features between the two considered frames, and undertaking, on the basis of said estimation, a long term motion analysis to obtain motion descriptors corresponding to the estimated camera motion parameters. Application: implementation of descriptors within MPEG–7.

Camera motion parameters estimation method.

The invention relates to a camera motion parameters estimation method, said parameters being intended to become descriptors in the MPEG-7 video indexing framework.

The last decades have seen the development of large databases of information, accessible to many people. These databases are composed of several types of media such as

5   text, images, sound, etc... The characterization, representation, indexing, storage, transmission and retrieval of such information constitute important issues in the usefulness of this technology. Whatever the level of sub-division at which video indexing can be contemplated, each information sub-division can be then indexed according to several criteria such as semantic information content, scene attributes, camera motion parameters, and so on. MPEG-

10   7, also named "Multimedia Content Description Interface" and focussing on content-based retrieval problems, will standardize generic ways to describe such multimedia content, using descriptors and description schemes that would be associated to multimedia material, in order to allow fast and efficient retrieval based on various types of features, such as text, color, texture, motion and semantic content. This standard will address applications that can be either

15   stored (on-line or off-line) or streamed (e.g. broadcast or video in the internet) and can operate in both real time and non-real time environments.

A schematic block diagram of a possible MPEG-7 processing chain, shown in Fig.1 and provided for processing any multimedia content, includes at the coding side a feature extraction sub-assembly 11 operating on said content, a normative sub-assembly 12,

20   including a module 121 for yielding the MPEG-7 definition language and a module 122 for defining the MPEG-7 descriptors and description schemes, a standard description sub-assembly 13, and a coding sub-assembly 14. The scope of the MPEG-7 standard is the sub-assembly 12, and the invention is located in the sub-assemblies 12 and 13. Fig.1 also shows the decoding side, including a decoding sub-assembly 16 (just after a transmission of the

25   coded data, or a reading operation of these stored coded data), and a search engine 17 working in reply to the actions controlled by the user.

In the MPEG-7 framework, efficient tools must be developed for many subjects like scene analysis or motion analysis, and particularly methods for camera motion feature

extraction. For a motion representation, two solutions can be proposed as a possible basis for global motion descriptors extraction : the perspective model, and the block matching method. The former is well suited for camera global motion, but cannot represent tridimensional translations, that have to be distinctly described each time it is possible.

5   Block matching motion compensation is used as a part of the predictive coding process that is widely used in video transmission for reducing the amount of information needed to encode a video sequence. Indeed, only a little fraction of an image changes from a frame to the following one, allowing a straightforward prediction from said previous frame. More precisely, each frame (i+1) is divided into a fixed number of blocks (usually square). For

10   each block (typically of 8 x 8 pixels), a search is made for the most similar block in a previous reference frame (i), over a predetermined area. The search criterion is generally the search of the best matching block, giving the least prediction error, usually computed as the mean absolute difference (which is easier to compute than for instance the mean square difference). For each block (in the present example, of 8 X 8 pixels) located in (x,y), the predicted image

15   (i+1) is then computed from image (i) according to the relation (1) :

$$B(i+1)[x,y] = B(i)[x-dx, y-dy] \quad (1)$$

with $(dx, dy) = \bar{v}$ = motion vector leading from $B(i)$, in the image (i), to $B(i+1)$, in the image (i+1).

          When starting from block matching motion vectors in order to estimate camera

20   movements, the main problem is then that the efficiency of the estimator of these vectors is only measured in terms of a coding criterion. Motion vectors do not necessarily correspond to the real motion of the scene. For example, in an area of homogeneous texture in the scene, the estimator could choose any of the blocks inside the texture, even if the motion vector is not representative of the global motion. However, although block matching represents a motion

25   that is not always consistent, this method will be preferred, because translations have to be distinctly described each time it is possible and the perspective model is not able to do it. Starting from motion vectors thus determined, some camera parameters will then be defined. Before describing the corresponding definition method, the camera model used in the present description is first presented.

30   A monocular camera moving through a static environment is considered. As can be seen in Fig.2, let 0 be the optical centre of the camera and OXYZ an external coordinates system that is fixed with respect to the camera, OZ being the optical axis. Let T_x, T_y, T_z be the translational velocity of OXYZ relative to the scene and R_x, R_y, R_z its angular

velocity. If (X,Y,Z) are the instantaneous coordinates of a point P in the tridimensional scene, the velocity components of P will be :

$$\overline{X} = -T_x - R_y.Z + R_z.Y \qquad (2)$$

$$\overline{Y} = -T_y - R_z.X + R_x.Z \qquad (3)$$

$$\overline{Z} = -T_z - R_x.Y + R_y.X \qquad (4)$$

The image position of P, namely p, is given in the image plane by the relation (5) :

$$(x,y) = \text{internal coordinates} = (f\frac{X}{Z}, f\frac{Y}{Z}) \qquad (5)$$

(where f is the focal length of the camera), and will move across the image plane with an induced velocity :

$$(u_x, u_y) = (\overline{x}, \overline{y}) \qquad (6)$$

After some computations and substitutions, the following relations are obtained :

$$u_x = f.\frac{\overline{X}}{Z} - f.\frac{X.\overline{Z}}{Z^2} \qquad (7)$$

$$u_x = \frac{f}{Z}(-T_x - R_y.Z + R_z.Y) - \frac{f.X}{Z^2}(-T_z - R_x.Y + R_y.X) \qquad (8)$$

and :

$$u_y = f.\frac{\overline{Y}}{Z} - f.\frac{Y.\overline{Z}}{Z^2} \qquad (9)$$

$$u_y = \frac{f}{Z}(-T_y - R_z.X + R_x.Z) - \frac{f.Y}{Z^2}(-T_z - R_x.Y + R_y.X) \qquad (10)$$

which can also be written :

$$u_x(x,y) = -\frac{f}{Z}.(T_x - x.T_z) + \frac{x.y}{f}.R_x - f(1+\frac{x^2}{f^2})R_y + y.R_z \qquad (11)$$

$$u_y(x,y) = -\frac{f}{Z}.(T_y - y.T_z) - \frac{x.y}{f}.R_y + f(1+\frac{y^2}{f^2})R_x - x.R_z \qquad (12)$$

Moreover, in order to include the zoom in the camera model, it is assumed that a zoom can be approximated by a single magnification in the angular domain. Such an hypothesis is valid if the distance of the nearest object in the scene is large compared to the change of focal length used to produce the zoom, which is usually the case.

A pure zoom is considered in Fig.3. Given a point located in the image plane, on (x,y) at a time t and on (x', y') at the next time t', the image velocity $u_x = x'-x$ along x

induced by the zoom can be expressed as a function of R_{zoom} (R_{zoom} being defined by the relation $(\theta' - \theta)/\theta$, as indicated in Fig.3), as shown below.

One has indeed : $\tan(\theta') = x'/f$ and $\tan(\theta) = x/f$, which leads to :

$$u_x = x' - x = [\tan(\theta') - \tan(\theta)].f \qquad (13)$$

The expression of $\tan(\theta')$ can be written :

$$\tan(\theta') = \tan[(\theta' - \theta) + \theta] = \frac{\tan(\theta' - \theta) + \tan(\theta)}{1 - \tan(\theta).\tan(\theta' - \theta)} \qquad (14)$$

Assuming then that the angular difference $(\theta' - \theta)$ is small, i.e. $\tan(\theta' - \theta)$ can be approximated by $(\theta' - \theta)$, and that $(\theta' - \theta).\tan\theta << 1$, one obtains :

$$u_x = x' - x = f. \left[\frac{(\theta' - \theta) + \tan(\theta)}{1 - (\theta' - \theta).\tan\theta} - \tan\theta \right] \qquad (15)$$

$$u_x = f. \frac{(\theta' - \theta).(1 + \tan^2(\theta))}{1 - (\theta' - \theta).\tan\theta} \qquad (16)$$

$$u_x = f.\theta.R_{ZOOM}. \frac{1 + \tan^2(\theta)}{1 - (\theta' - \theta).\tan\theta} \qquad (17)$$

which is practically equivalent to :

$$u_x = x' - x = f.\theta.R_{zoom}.(1 + \tan^2\theta) \qquad (18)$$

This result can be rewritten :

$$u_x = f.\tan^{-1}\left(\frac{x}{f}\right).R_{zoom}.\left(1 + \frac{x^2}{f^2}\right) \qquad (19)$$

and, similarly, u_y is given by :

$$u_Y = f.\tan^{-1}\left(\frac{y}{f}\right).R_{zoom}.\left(1 + \frac{y^2}{f^2}\right) \qquad (20)$$

The velocity $u = (u_x, u_y)$ corresponds to the motion induced in the image plane by a single zoom. A general model in which all the rotations, translations (along X and Y axis) and zoom are taken into account can then logically be defined.

This general model can be written as the sum of a rotational velocity, representing rotational and zoom motions, and a translational velocity, representing the X and Y translations (i.e. tracking and booming respectively)

$$\begin{cases} u_x = u_x^{trans} + u_x^{rot} \\ u_y = u_y^{trans} + u_y^{rot} \end{cases} \qquad (21)$$

with :

$$\left\{ \begin{array}{l} u_x^{trans} = -\dfrac{f}{Z}.T_x \\[2ex] u_y^{trans} = -\dfrac{f}{Z}.Ty \end{array} \right\}$$

$$\left\{ \begin{array}{l} u_x^{rot} = \dfrac{x.y}{f}.R_x - f.(1+\dfrac{x^2}{f^2}).R_y + y.R_z + f.\tan^{-1}(\dfrac{x}{f}).(1+\dfrac{x^2}{f^2}).R_{zoom} \\[3ex] u_y^{rot} = -\dfrac{x.y}{f}.R_y + f.(1+\dfrac{y^2}{f^2}).R_x - x.R_z + f.\tan^{-1}(\dfrac{y}{f}).(1+\dfrac{y^2}{f^2}).R_{zoom} \end{array} \right\}$$

equations in which only translational terms depend on the object distance Z.

The article "Qualitative estimation of camera motion parameters from video

5   sequences", by M.V. Srinivasan and al., Pattern Recognition, vol.30, n°4, 1997, pp.593-605, describes for extracting camera motion parameters from a sequence of images a technique using the camera equations (21) to (23). More precisely, the basic principle of this technique is explained in part 3 (pp.595-597) of said article. The technique, implemented by finding the best values of R_x, R_y, R_z and R_{zoom} that create a flow field which, when subtracted from the

10   original optic flow field, then results in a residual flow field wherein all the vectors are parallel, uses an iterative method minimizing deviations from parallelism of the residual flow vectors, by means of an advantageous sector-based criterion.

At each step of this iterative method, the optic flow due to the current camera motion parameters is calculated according to one of two different camera models. A first

15   model assumes that the angular size of the visual field (or the focal length f) is known : this means that the ratios x/f and y/f in the equations (23) can be calculated for each point in the image, said equations then allowing to calculate the optic flow exactly.

A second model assumes that the visual field of the camera is not known. Small field approximations (x/f and y/f very lower than 1) are then necessary before applying the

20   equation (23), which leads to the equations (24) and (25) :

$$u_x^{rot} \square - f.R_y + y.R_z + x.R_{zoom} \qquad\qquad (24)$$

$$u_y^{rot} \square - f.R_x - x.R_z + y.R_{zoom} \qquad\qquad (25)$$

It appears that the first model, which is the one taking into account panning or tilting distorsions, produces more accurate results when the visual field of the camera is large

25   and known. Unfortunately, the focal length is generally not known, which leads to use the second model, only on a restricted area of the image when the visual field is suspected to be

large. However, this second model is not a satisfying solution since it is not possible to distinguish panning from tracking.

5   While horizontal and vertical trackings produce flow fields in which all vectors are truly parallel, this is not the case with pan and tilt unless the visual field of the camera is small, for instance : 20° x 20° (large visual fields lead to distorsions in the velocity fields arising from the planar geometry of the imaging surface). The flow field produced by a zoom is also distorted (far away from the centre, motion vectors are no longer radially oriented). If

10   the visual field is large, the resulting distorsions can be used to enable translationel motions to be uniquely distinguished from pan and tilt.

It is therefore an object of the invention to improve the scheme for camera motion features estimation from motion vectors by proposing a camera motion parameters estimation method that is able, each time it is physically possible, to make the difference

15   between tracking and panning and differently carry out them when the visual field is large, even if it is unknown, but without making this difference when the visual field is small.

To this end, the invention relates to an estimation method provided for considering a sequence of successive video frames subdivided into blocks and processing this sequence, wherein said processing operation comprises according to the successive steps of :

20   - extracting from said video sequence vectors corresponding to the motion between two successive frames, said motion vectors forming the camera velocity field ;

- preprocessing the camera velocity field, in order to reduce the amount of data and the heterogeneousness of said extracted motion vectors ;

- estimating for each pair of frames, from said preprocessed field, camera

25   features between the two considered frames ;

- undertaking on the basis of said estimation a long term motion analysis to obtain motion descriptors corresponding to the estimated camera motion parameters.

The main idea is that if a large visual field generates distorsions in the velocity field in most cases, the same distorsions should also be useful. In other words, if the focal

30   length (or the visual field, which is the same information with an image size scale factor) is included in the minimization process that uses the first model hereinabove mentioned, it should be correctly estimated when the visual field is not too small and when there is actually one of the zoom, pan, tilt or roll motion components (which represents an important part of

real cases ; this focal length estimation would not be meaningful if the visual field is too small or if there is only a tracking motion).

5   The present invention will now be described, by way of example, with reference to the accompanying drawings in which :

- Fig.1 is a schematic block diagram of a possible MPEG-7 processing chain ;

- Fig.2 illustrates for a camera an external coordinates system 0XYZ moving with the camera and shows, for a focal length f, the retinal coordinates (x,y) corresponding to a

10   point P in the tridimensional scene ;

- Fig.3 illustrates a zoom model included in the camera model ;

- Fig.4 gives an overview of a complete camera motion analysis system carrying out the estimation method according to the invention ;

- Fig.5 illustrates the filtering technique used in the system according to the

15   invention ;

- Fig.6 illustrates how, when implementing the downhill simplex method, a new simplex is constructed ;

- Fig.7 shows a flowchart illustrating the downhill simplex algorithm used for the parallelism cost function.

20

          A complete scheme of an implementation of the estimation method according to the invention is illustrated in Fig.4. It may be noted that, since MPEG-7 will be a multimedia content description standard, it does not specify a particular coding type. A

25   process of descriptors formation must therefore work on all types of coded data (either compressed or uncompressed). Nevertheless, as most of the video data obtained from the input frames are generally available in the MPEG format, (there are therefore compressed), it is advantageous to use directly the motion vectors provided by the MPEG motion compensation. If the video data are available in the uncompressed domain, a block-matching method is

30   therefore implemented in a motion vector generation device 41.

          Whatever the case, once motion vectors have been read or extracted from the video sequence (between two successive frames), a downsampling and filtering device 42 is provided. The goal of the filtering of the velocity field is to reduce the amount of data and the heterogeneousness of the motion vectors. This operation is followed by an instantaneous

estimation, in a device 43, of the camera features. This estimation is based on the method previously described, improved in order to use motion vectors and to include the focal length in the minimization process, and leads to one features vector for each pair of frames. The set of features vectors within the whole considered sequence is then finally received by a long-term motion analysis device 44 that outputs motion descriptors which may be used to index the sequence in terms of camera motion in a content-based retrieval context, especially in the MPEG-7 video indexing framework.

Two main problems justify the preprocessing step in the device 42 : the heterogeneousness of the motion vectors, above all in the low-frequency parts of the image or where texture is very homogeneous, and the too small size of the blocks. The downsampling and filtering process is provided for reducing the amount of vectors by downsampling the original field and simultaneously rejecting the vectors that are not consistant according to the global information. A confidence mask, calculated for each vector, is used : it is a criterion varying between 0 and 1 according to the level of confidence of each motion vector and allowing to decide if the vectors are taken into account or not. An example of confidence mask may be to consider that for any theoretical camera motion, a motion vector cannot vary too much : close vectors have close values. One can then measure a confidence level according to the distance from each vector to its neighbourhood, which can be for instance represented by its average value or, preferably, its median (because it is less sensitive to big isolated errors). The confidence mask $C_{i,j}$ is therefore defined by the equation (26) :

$$C_{i,j} = e^{-\left\| \bar{v}_{i,j} - \bar{v}_{median} \right\|^2} \qquad (26)$$

Fig.5 illustrates the filtering technique : the filtered field (right side) contains four times as fewer blocks as the original field (left side). The vector representing the motion of a new block is computed according to the motion vectors of the four original blocks, and their confidence level is calculated according to the neighbourhood as indicated. The motion vector for the new block is the weighted mean of its old smaller blocks :

$$\bar{V}_{m,n}(filt) = \frac{\displaystyle\sum_{i=2(m-1)+1}^{2(m-1)+2} \sum_{j=2(n-1)+1}^{2(n-1)+2} C_{i,j} \cdot \bar{V}_{i,j}}{\displaystyle\sum_{i=2(m-1)+1}^{2(m-1)+2} \sum_{j=2(n-1)+1}^{2(n-1)+2}} \qquad (27)$$

The device 43, provided for computing for each pair of frames, from the filtered motion vectors field, a feature vector that contains the camera motion information between the two considered frames, implements an estimation algorithm which is now detailed.

First, the confidence mask is computed, from the equation (26). Then the parallelisation process starts. Each time a motion vector is taken into account in the computation of the cost function or of the resultant vector, it is weighted by its confidence mask. The following equations then allow to compute the best values of R_x, R_y, R_z, R_{zoom} and the focal length f that give a residual field in which all the vectors are parallel :

$$\bar{R}^{estim} = \left[\hat{R}_x, \hat{R}_y, \hat{R}_z, R_{zoom}, \hat{f}\right] = \arg\min\left\{P(\hat{R})\right\} \qquad (28)$$

$$\text{where } P(\bar{R}) = \sum_i \sum_j \left\|\bar{v}_{i,j}^{residual}(\bar{R})\right\|^2 . \theta_{i,j} . C_{i,j} \qquad (29)$$

with :

$$\bar{v}_{i,j}^{residual}(\bar{R}) = \bar{v}_{i,j} - \begin{bmatrix} u_x^{rot}(\bar{R}) \\ u_y^{rot}(\bar{R}) \end{bmatrix} \qquad (30)$$

$$\text{and } \theta_{i,j} = \text{angle}(\bar{v}_{i,j}^{residual}, \bar{V}^{residual}),$$

$$\bar{V}^{residual} = \frac{\sum_i \sum_j \bar{v}_{i,j}^{residual} . C_{i,j}}{\sum_i \sum_j C_{i,j}} \qquad (31)$$

In the case of a non-translational motion in a large visual field, the residual vectors would not be parallel but should ideally be close to zero. This remark leads to compute the β ratio given by the equation (32) :

$$\beta = \frac{\left\|\sum \bar{v}_{i,j}^{residual}(\bar{R}^{estim})\right\|}{\sum \left\|\bar{v}_{i,j}^{residual}(\bar{R}^{estim})\right\|} \qquad (32)$$

which indicates the parallelism of the residual field. This is the ratio of the magnitude of the resultant of the residual flow vectors to the sum of the magnitudes of the residual flow vectors : β = 1 implies that the residual vectors are perfectly aligned, while β = 0 implies that the residual vectors are randomly oriented with respect to each other. Moreover, to check the presence of a significant tracking component in the camera motion, the strength of the residual flow field is compared to that of the original flow field by computing the following ratio α, given by the equation (33) :

$$a = \frac{\text{mean}^{(*)}\left(\left\|\bar{v}_{i,j}^{residual}(\bar{R}^{estim})\right\|\right)}{\text{mean}^{(*)}\left(\left\|\bar{v}_{i,j}\right\|\right)} \qquad (33)$$

The "mean(*)" operator represents the weighted mean of its arguments, according to the confidence mask. These two ratios allow to check for the presence and the amount of tracking components as shown below :

A) if $\beta \sim 0$, no tracking motion ;

B) if $\beta \sim 1$:

if $\alpha \sim 0$, negligible tracking motion ;

if $\alpha \sim 1$, significant tracking motion :

$$\hat{T}_x = -V_x^{residual}$$

$$\hat{T}_y = -V_y^{residual}$$

These ratios also give an idea of the results relevance.

It must be noted that the estimated components of translational motion, namely $\hat{T}_x$ and $\hat{T}_y$ do not represent exact components of the first model, but a weighted mean within the whole image of $f.\dfrac{T_x}{z}$ and $f.\dfrac{T_y}{z}$, since the depth of each block is not known. However, they are good representations of apparent tracking motion in the image.

For the parallelism cost function, the downhill simplex minimization algorithm has been chosen because it does not involve computation of the derivatives, which would be impossible here since the original field is part of the cost function. The downhill simplex method is a multidimensional method that requires only function evaluations, not derivatives, unlike many others methods like all the gradient-based ones. It is an essential point here since there is no analytical form of the cost function. A simplex is the geometrical figure consisting, in N dimensions, of N+1 points (or vertices) and all their interconnecting line segments, polygonal faces, etc... : in two dimensions, a simplex is a triangle, in three dimensions it is a tetrahedron, and so on. In general, only simplexes that are non degenerate (i.e. that enclose a finite n-dimensional volume) are considered. If any point of a simplex is taken as the origin, the other N points define vector directions that span the N-dimensional vector space. In one-dimensional minimization, it is possible to bracket a minimum (dichotomy) so that the success of isolation is guaranteed.

Even if there is no analogous procedure in multidimensional space, the downhill simplex method is the one that best fits this idea. The algorithm starts with an N-vector of independent variables as the first point to try and is then supposed to make its own way downhill through the space until it encounters a minimum to enclose, at least a local one. The method must be started not just with a single point, but with N+1 points, defining an

initial simplex. Then, at each step, a new simplex is constructed. As shown in Fig.6 (left part) most steps move the point of the simplex where the function is largest ("highest point") through the opposite face of the simplex to a lower point (reflections). These kinds of steps conserve the volume of the simplex. The procedures for expansion and contraction enable the simplex method to accelerate along a successfull track of improvement and to home in on the optimum conditions. The algorithm stops when the required objectives are reached, practically when the decrease in the function value is fractionally smaller than some tolerance ε .

The used complete algorithm is illustrated thanks to the flowchart given in Fig.7. After initialization (START operation) and ranking by function value (step 62), tests $(f(.)<f(.) ?)$ 63, 64, 65, 66, 67, 68 are carried out :

1) $f(R) < f(W') ?$

    a) if yes (Y), $f(R) < f(B') ?$

        i) if yes (Y), $f(E) < f(B') ?$

          - if yes (Y), replace W by E (output 71) ;

          - if no (N), replace W by R (output 72) ;

        ii) if no (N), replace W by R ;

    b) if no (N), $f(R) < f(W) ?$

        i) if yes (Y), replace W by R ;

        ii) if no (N), C is defined (COMPUTE C) by the expression (34) :

$$C = \arg\min [f(x)] \text{ pour } x \, \varepsilon \, \{C^{(+)}, C^{(-)}\} \qquad (34)$$

and a further test is carried out :

2) $f(C) < f(W) ?$

    (a) if yes (Y), replace W by C (output 73)

    (b) if no (N), make multiple contraction (output 74) ;

3) after the outputs 71 to 74, $f(W) - f(B) < \varepsilon ?$

    - if no (N), feedback connection towards the step 62 ;

    - if yes (Y), end of the algorithm (STOP).

with the following notations :

    - $f = R^N \rightarrow R$, the function to minimize (here a parallelism function) ;

    - W, $C^{(-)}$, $C^{(+)}$, R, E, the N-dimensions points ($C^{(-)}$ = negative contraction, $C^{(+)}$ = positive contraction, R = reflection, E = expansion, as illustrated in Fig.6, right part)

    - W', B, B' the antecedents of the second worst, the best and the second best values of f (then $f(B) < f(B') <...< f(W') < f(W)$) ;

The possible destination points are computed as given by :

$$\begin{cases} R = 2M - W \\ E = M + \gamma \ (M - W) \\ C^{(+)} = M + \beta^{+} (M - W) \\ C^{(-)} = M - \beta^{-} (M - W) \end{cases}$$

Since their influence on the optimization performances is difficult to handle, the coefficients have been set to their simplest value, i.e : $\gamma = 2$, $\beta^{(+)} = \beta^{(-)} = 0,5$.

CLAIMS:

1.   A camera motion parameters estimation method provided for considering a sequence of successive video frames subdivided into blocks and processing this sequence, wherein said processing operation comprises the successive steps of :

- extracting from said video sequence vectors corresponding to the motion between two successive frames, said motion vectors forming the camera velocity field ;

- preprocessing the camera velocity field, in order to reduce the amount of data and the heterogeneousness of said extracted motion vectors ;

- estimating for each pair of frames, from said preprocessed field, camera features between the two considered frames ;

- undertaking on the basis of said estimation a long term motion analysis to obtain motion descriptors corresponding to the estimated camera motion parameters.

2.   A method according to claim 1, wherein said preprocessing step comprises, in order to reduce the amount of vectors, the sub-steps of :

- downsampling said original velocity field of motion vectors ;

- simultaneously, rejecting the vectors that are not consistent according to the global information.

3.   A method according to claim 2, wherein said preprocessing sub-step includes the computing operation of a vector representing the motion of a new block according to the motion vectors of the four original blocks and of an associated confidence level.

4.   A method according to claim 3, wherein said confidence level is given by the distance from each vector to its neighbourhood.

5.   A method according to claim 4, wherein said neighbourhood is represented by a value $\bar{V}(m)$ such as the mean value or the median, leading for the confidence mask to a value $C_{i,j}$ having an expression of the type :

$$C_{i,j} = e^{-\left\|\vec{v}_{i,j} - \bar{v}(m)\right\|^2}$$

where (i,j) defines the original block and $\bar{V}_{i,j}$ its original motion vector.
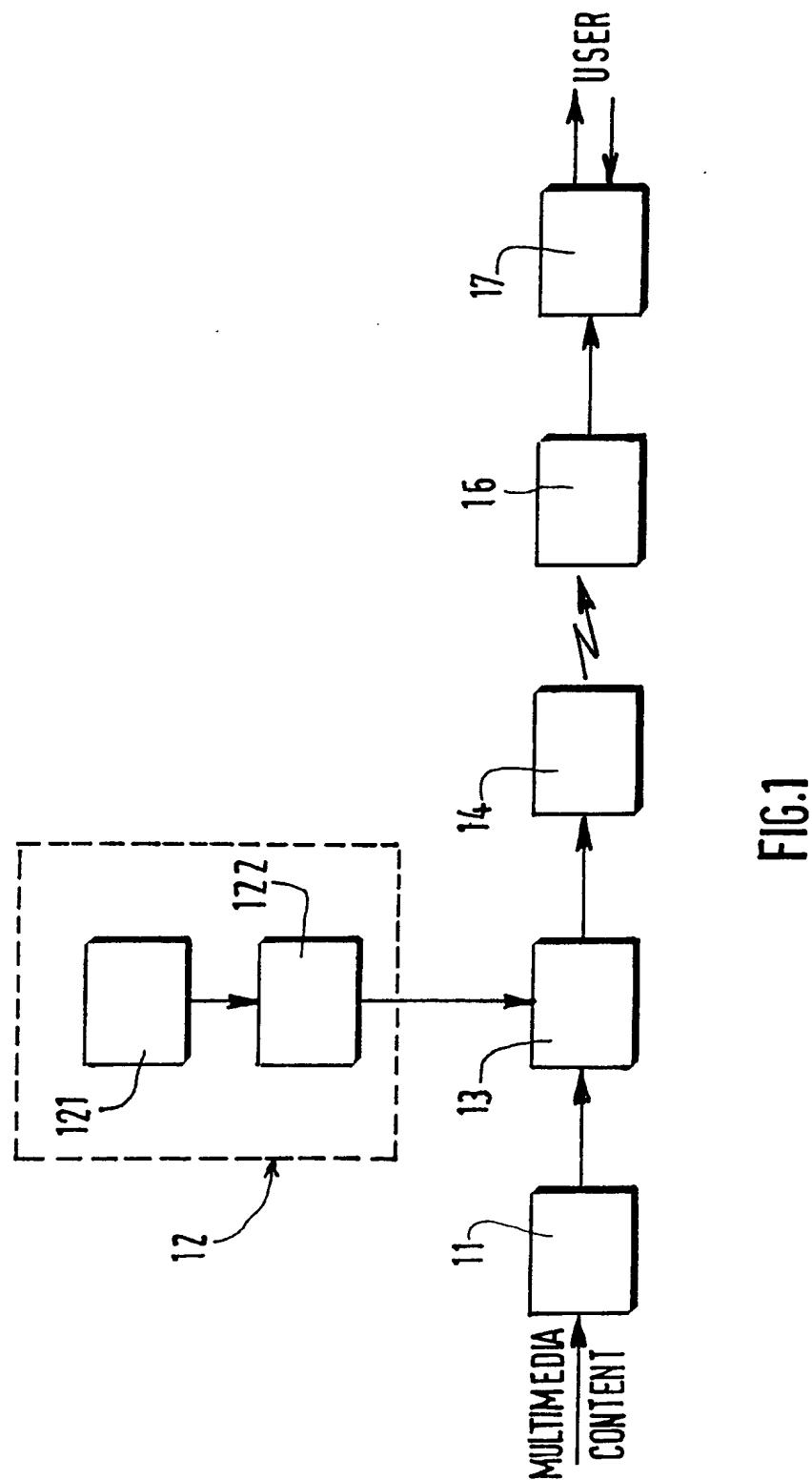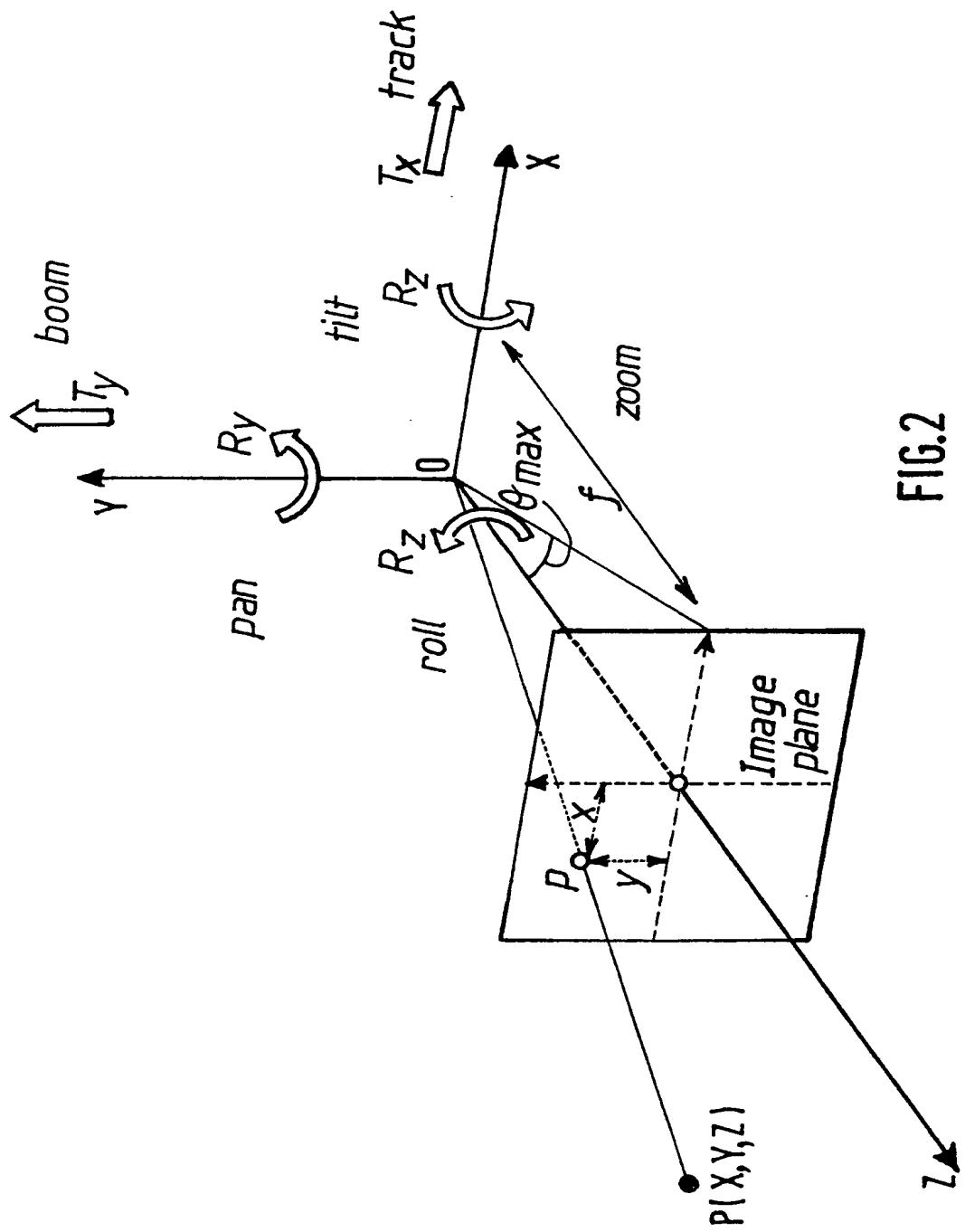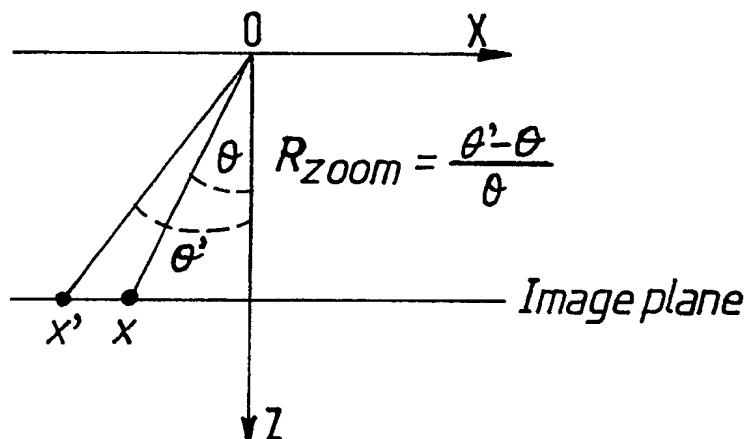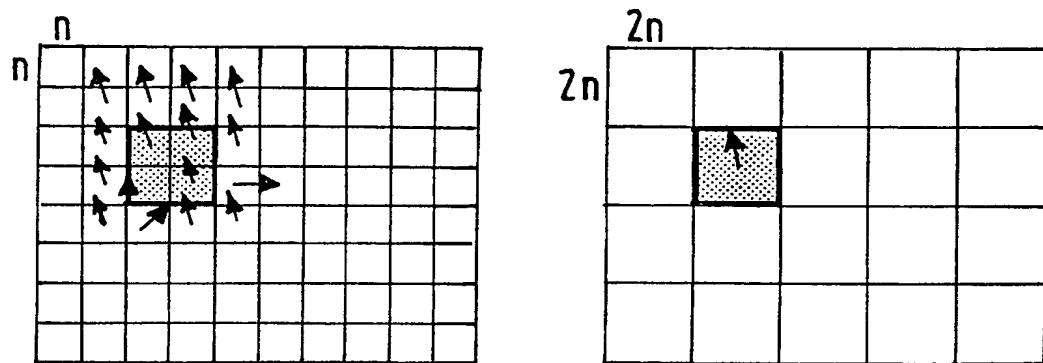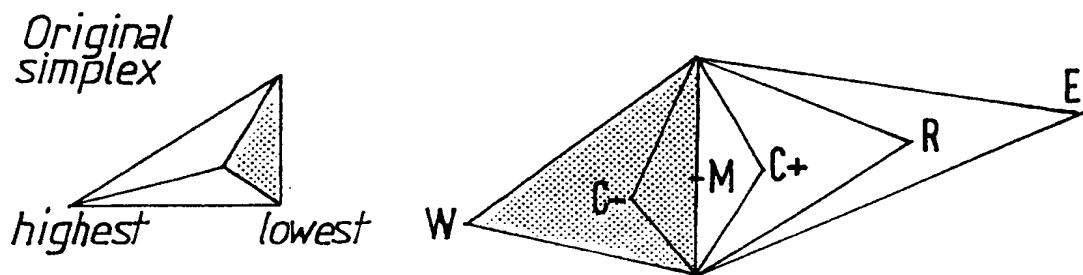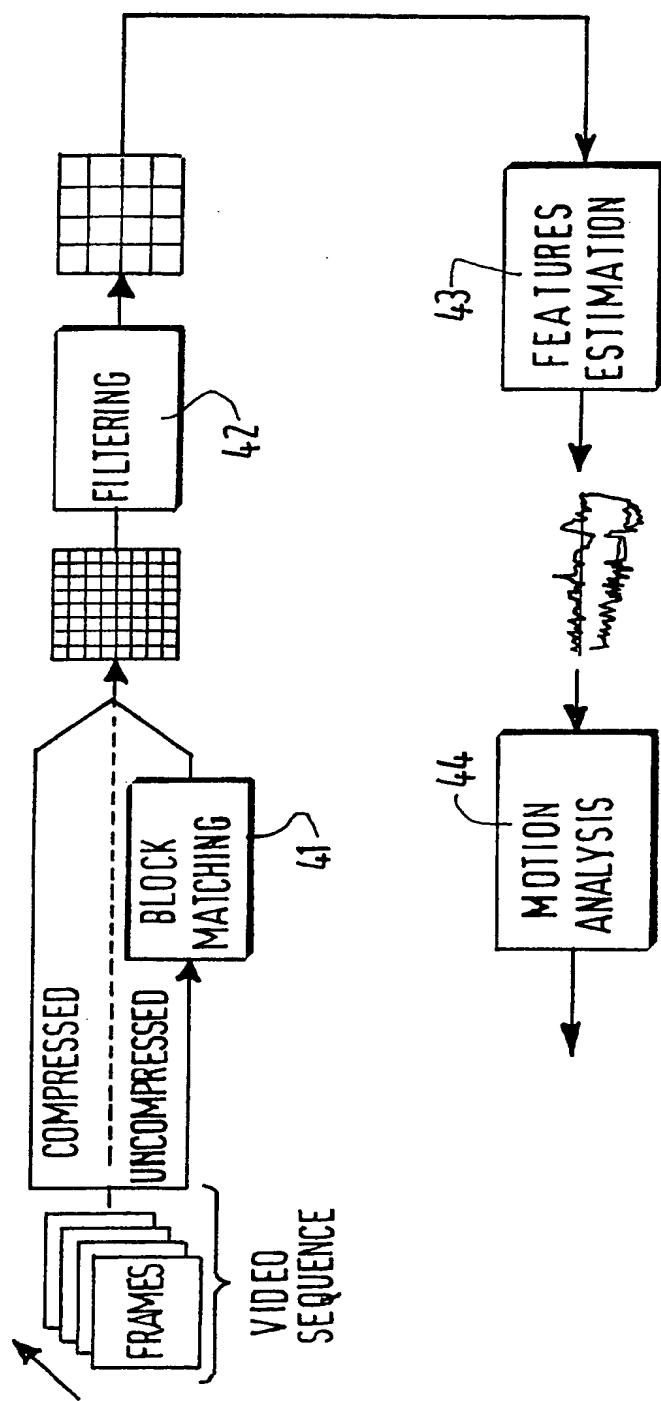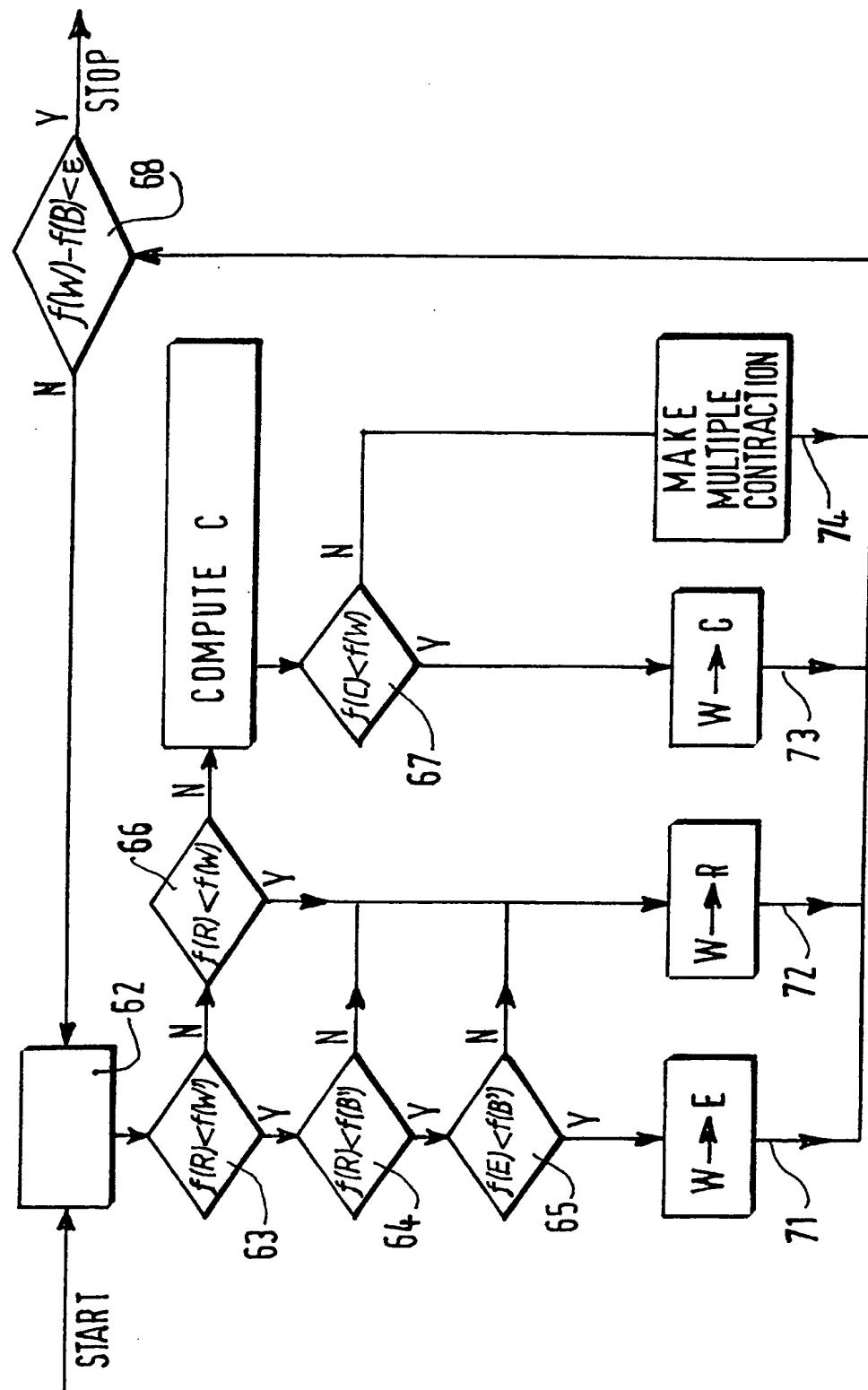
FIG.1

FIG.2

$$R_{zoom} = \frac{\theta' - \theta}{\theta}$$

Image plane

FIG.3

FIG.5

Original simplex

highest   lowest

FIG.6

FIG.4

FIG.7

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7   H04N7/26   G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7   H04N   G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SRINIVASAN M V ET AL: "QUALITATIVE ESTIMATION OF CAMERA MOTION PARAMETERS FROM VIDEO SEQUENCES" PATTERN RECOGNITION,US,PERGAMON PRESS INC. ELMSFORD, N.Y, vol. 30, no. 4, 1 April 1997 (1997-04-01), pages 593-605, XP000690821 ISSN: 0031-3203 cited in the application page 595, left-hand column, paragraph 3 -page 601, left-hand column, paragraph 4.3 --- -/--	1-5

[X] Further documents are listed in the continuation of box C.   [X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 May 2000	30/05/2000

Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Authorized officer Schoeyer, M

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	KIM E T ET AL: "EFFICIENT LINEAR THREE-DIMENSIONAL CAMERA MOTION ESTIMATION METHOD WITH APPLICATIONS TO VIDEO CODING" OPTICAL ENGINEERING,US,SOC. OF PHOTO-OPTICAL INSTRUMENTATION ENGINEERS. BELLINGHAM, vol. 37, no. 3, 1 March 1998 (1998-03-01), pages 1065-1077, XP000771086 ISSN: 0091-3286 page 1066, left-hand column, paragraph 2 -page 1071, right-hand column, paragraph 4.4	1-5
A	US 5 635 994 A (KEESEN HEINZ-WERNER ET AL) 3 June 1997 (1997-06-03) column 1, line 23 -column 1, line 31	1-5
A	US 5 259 040 A (HANNA KEITH J) 2 November 1993 (1993-11-02) abstract	1-5
A	EP 0 509 208 A (INST PERSONALIZED INFORMATION) 21 October 1992 (1992-10-21) column 3, line 42 -column 3, line 53	1-3
A	OOMOTO, EITESTSU ET. AL.: "Ovid: Design and Implementation of a Video-Object Database System" IEEE TRANSACTIONS OF KNOWLEDGE AND DATA ENGINEERING, vol. 5, no. 4, 4 August 1993 (1993-08-04), pages 629-643, XP002134326 abstract	1-5

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 5635994	A	03-06-1997	DE 4342305 A		29-06-1995
			CN 1117694 A		28-02-1996
			DE 59408390 D		15-07-1999
			EP 0658056 A		14-06-1995
			JP 7203451 A		04-08-1995
			SG 49199 A		18-05-1998
US 5259040	A	02-11-1993	EP 0606385 A		20-07-1994
			JP 6511339 T		15-12-1994
			WO 9307585 A		15-04-1993
EP 0509208	A	21-10-1992	JP 2677312 B		17-11-1997
			JP 4281679 A		07-10-1992
			DE 69219944 D		03-07-1997
			DE 69219944 T		05-02-1998
			US 5267034 A		30-11-1993